



High-quality sika deer omics data and integrative analysis reveal genic and cellular regulation of antler regeneration

Zihe Li, Ziyu Xu, Lei Zhu, et al.



Genome Res. published online November 14, 2024

Access the most recent version at doi:[10.1101/gr.279448.124](https://doi.org/10.1101/gr.279448.124)

P<P	Published online November 14, 2024 in advance of the print journal.
Accepted Manuscript	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
Open Access	Freely available online through the <i>Genome Research</i> Open Access option.
Creative Commons License	This manuscript is Open Access. This article, published in <i>Genome Research</i> , is available under a Creative Commons License (Attribution-NonCommercial 4.0 International license), as described at http://creativecommons.org/licenses/by-nc/4.0/ .
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .

CRISPR and RNAi Genetic Screening.
Your new superpower.

LEARN MORE



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Published by Cold Spring Harbor Laboratory Press

1 **High-quality sika deer omics data and integrative analysis reveal genic and cellular**
2 **regulation of antler regeneration.**

3 Zihe Li^{1,11}, Ziyu Xu^{2,3,11}, Lei Zhu^{6,7,11}, Tao Qin^{1,11}, Jinrui Ma^{1,11}, Zhanying Feng^{2,10,11}, Huishan
4 Yue¹, Qing Guan⁹, Botong Zhou¹, Ge Han¹, Guokun Zhang⁸, Chunyi Li⁸, Shuaijun Jia^{6,7},
5 Qiang Qiu^{1*}, Dingjun Hao^{6,7*}, Yong Wang^{2,3,4,5*}, Wen Wang^{1,9*}

6

7

8 ¹ New Cornerstone Science Laboratory, Shaanxi Key Laboratory of Qinling Ecological

9 Intelligent Monitoring and Protection, School of Ecology and Environment, Northwestern

10 Polytechnical University, Xi'an 710072, China.

11 ²CEMS, NCMIS, HCMS, MADIS, Academy of Mathematics and Systems Science, Chinese

12 Academy of Sciences, Beijing 100190, China.

13 ³School of Mathematics, University of Chinese Academy of Sciences, Chinese Academy of

14 Sciences, Beijing 100049, China.

15 ⁴Key Laboratory of Systems Health Science of Zhejiang Province, School of Life Science,

16 Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences,

17 Hangzhou, 310024, China.

18 ⁵Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences,

19 Kunming 650223, China.

20 ⁶Department of Spine Surgery, Honghui Hospital, Xi'an Jiaotong University, Xi'an, Shaanxi

21 710054, China

22 ⁷Shaanxi Key Laboratory of Spine Bionic Treatment, Xi'an, Shaanxi 710054, China

23 ⁸ Institute of Antler Science and Product Technology, Changchun Sci-Tech University;

24 130600 Changchun, China

25 ⁹Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology,

26 Chinese Academy of Sciences, Kunming, Yunnan 650223, China.

27 ¹⁰Department of Statistics, Department of Biomedical Data Science, Bio-X Program, Stanford

28 University, Stanford, CA 94305, USA

29 ¹¹These authors contributed equally: Zihe Li, Ziyu Xu, Lei Zhu Tao Qin, Jinrui Ma, Zhanying

30 Feng

31

32 * Correspondent authors: wenwang@nwpu.edu.cn (W.W.), ywang@amss.ac.cn (Y.W.),

33 haodingjun@mail.xjtu.edu.cn (D.J.H.), qiuqiang@nwpu.edu.cn (Q.Q.)

34

35 **Abstract**

36 Antler is the only organ that can fully regenerate annually in mammals. However, the
37 regulatory pattern and mechanism of gene expression and cell differentiation during this
38 process remain largely unknown. Here, we obtain comprehensive assembly and gene
39 annotation of the sika deer (*Cervus nippon*) genome. Together with large-scale chromatin
40 accessibility and gene expression data, we construct gene regulatory networks involved in
41 antler regeneration, identifying four transcription factors, *MYC*, *KLF4*, *NFE2L2*, and *JDP2*
42 with high regulatory activity across whole regeneration process. Comparative studies and
43 luciferase reporter assay suggest the *MYC* expression driven by a cervid-specific regulatory
44 element might be important for antler regenerative ability. We further develop a model called
45 cTOP which integrates single-cell data with bulk regulatory networks and find *PRDMI*,
46 *FOSL1*, *BACH1*, and *NFATC1* as potential pivotal factors in antler stem cell activation and
47 osteogenic differentiation. Additionally, we uncover interactions within and between cell
48 programs and pathways during the regeneration process. These findings provide insights into
49 the gene and cell regulatory mechanisms of antler regeneration, particularly in stem cell
50 activation and differentiation.
51

52 **Introduction**

53 The cyclic renewal of deer antlers, the only known mammal organ that can fully
54 regenerate annually, has long captivated numerous biologists(Li et al. 2014). Each spring,
55 deer shed their hard antlers, and the pedicle scar heals as new bone and cartilage regenerate
56 from the pedicle periosteum. During late spring and early summer, antlers grow and calcify
57 over 3 to 4 months at a rate of 2.75 cm per day, potentially reaching a weight of up to 15 kg
58 by the end of the growth season (Price et al. 2005; Landete-Castillejos et al. 2019). In autumn,
59 antlers rapidly calcify and shed their enveloping velvet skin, leaving an exposed bony surface.
60 The cycle concludes with the casting of the antlers in spring, initiating a new regeneration
61 phase. Researchers have discovered that antler regeneration is driven by stem cells located in
62 the antler pedicle periosteum(Wang et al. 2019a). Casting of the pedicle periosteum results
63 failure of regeneration. Several somatic stem cell markers such as *NT5E*, *THY1* and *ENG*
64 have been found expressed in antler stem cells(Dong et al. 2020). Some researcher found
65 embryonic stem cell factors such as *MYC*, *KLF4* and *POU5F1* expressed in the regenerative
66 antler stem cell (Dąbrowska et al. 2016), while a single-cell transcriptome study suggested
67 that these factors were specific to antlerogenic stem cells(Ba et al. 2022). Our recent work
68 further unveiled the cell atlas of antler regeneration and elucidated the vital role of antler
69 blastema progenitor cells (ABPCs) differentiated from *PRRX1*⁺ mesenchymal stem cells
70 (PMCs) in the bone regeneration of sika deer antler(Qin et al. 2023). This study also
71 demonstrated the involvement of early developmental pathways, including the WNT, TGF- β ,

72 and FGF pathways, during antler regeneration. Despite these advancements, a comprehensive
73 understanding of the complex gene and cell regulatory dynamics in deer antler regeneration
74 remains elusive.

75 Gene regulatory networks (GRNs) constructed from transcriptomic and epigenomic
76 data have been widely used to explain complex biological processes such as development and
77 regeneration(Johnston et al. 2021; Jimenez et al. 2022). Recent studies on retinal regeneration
78 have identified key GRNs and regulatory factors, such as *NFIA/B*, which constrain
79 mammalian retinal regeneration(Hoang et al. 2020). However, most of these studies are
80 limited to model species due to the lack of necessary data resources, including high-quality
81 genomic data in non-model organisms. The quality of genomic data significantly impacts
82 multiomic studies(Chisanga et al. 2022). Among Cervidae, the most comprehensive genomic
83 data so far is for red deer (*Cervus elaphus*)(Pemberton et al. 2021), sequenced and assembled
84 by the Darwin Tree of Life Project and annotated using the latest RefSeq pipelines(Blaxter
85 2022; Pruitt et al. 2014). However, antler regeneration related data are limited and difficult to
86 obtain for red deer in contrast to sika deer (*Cervus nippon*), as many herds of sika deer are
87 reared for antler medicine production in China(Mccullough et al. 2009). Therefore, the sika
88 deer could serve as a valuable model for mammalian regeneration research due to the ease of
89 sampling. Several sika deer genome assemblies derived from short-read or long-read
90 sequencing are available(Han et al. 2022; Xing et al. 2023; Qin et al. 2023; Chen et al. 2019).
91 Most of these studies focused on evolutionary analysis, revealing the genetic basis of unique
92 traits such as high-tannin diet adaptation, rapid antler growth and cancer resistance in sika

93 deer. These resources have been utilized in transcriptomic research on deer antler(Zhang et al.
94 2022). However, comprehensive genomic data, including high-quality genome assembly and
95 annotation for sika deer, remain lacking. This gap, combined with the challenges of
96 determining regulatory relationships between regulatory elements (REs) and target genes
97 (TGs) using standard ATAC-seq and histone ChIP-seq methods, have greatly hindered efforts
98 to unravel the gene and cell regulatory mechanisms underlying antler regeneration.

99 Recently, we developed PECA2 (paired gene expression and chromatin accessibility) to
100 more accurately infer regulatory relationships by integrating paired chromatin accessibility
101 data and gene expression data in mouse and human(Duren et al. 2017). However, this
102 approach relies on extensive paired chromatin accessibility and gene expression data, which
103 are currently unavailable for deer species. In this work, we aim to generate comprehensive
104 omics data for sika deer, including genomic, transcriptomic, and chromatin accessibility data.
105 By leveraging high-quality data and newly developed methods, we seek to gain insights into
106 the multi-scaled regulatory dynamics of antler regeneration.

107 **Results**

108 **Genome sequencing, assembly and annotation for sika deer**

109 We first conducted flow cytometry to estimate the sika deer genome size as 3.5 Gb (Fig.
110 S1A-C). Next, we employed multiple sequencing technologies to achieve a high-quality and
111 highly contiguous genome assembly (Fig. 1A). Initially, we obtained 91.52 Gb (31x) of

112 PacBio HiFi reads and 360 Gb (120x) of Hi-C BGI reads (Supplementary Table1) from a
113 male sika deer. We utilized Hifiasm(Cheng et al. 2022) with these HiFi and Hi-C data, as well
114 as ONT long reads (N50 > 38k) from another individual in a previous study(Qin et al.
115 2023),to generate well-phased contig assemblies. Finally, we obtained two haplotype
116 assemblies with sizes of 3.1 Gb and 3.0 Gb, respectively. We then employed a two-step
117 haplotype-aware scaffolding strategy to finely phase and scaffold chromosomes for each
118 haplotype using the Hi-C data (see details in Methods). Finally, we anchored the contigs to 66
119 phased pseudochromosomes ($32 \times 2 + X, Y$) (Fig. 1B) (Supplementary Table 2). Compared to
120 previous assemblies for sika deer (Han et al. 2022; Xing et al. 2023), our assembled haplotype
121 sizes were much larger and closer to the estimated size. The majority of the additional
122 assembly originated from satellite-rich (47.48%) heterochromatin sequences (Supplementary
123 Table 2). The euchromatic contigs scaffolded to pseudochromosomes have a total size of
124 about 2.6 Gb, similar to other ruminant assemblies. Additionally, we found that the X and Y
125 Chromosomes in the previous genome assembly lacked the attachment of the PAR
126 (pseudoautosomal region) (Fig. S1 D-E). In our assembly, the PAR region was fully attached
127 to the X and Y Chromosomes and showed good collinearity with those of other mammals
128 (Fig. S1 F). Compared with cow, the distal end of PAR in sika deer extended to *SHROOM2*,
129 which is more similar to pig(Liu et al. 2019a). Seventeen chromosomes contain no more than
130 three contigs, indicating high continuity. Other quality metrics of our assembly further
131 demonstrated the high quality of our data (Supplementary Table 2).

132 We annotated 42.80% of repetitive sequences in the genome of sika deer. Notably, the

133 proportion of satellite sequences (9.96%), was much higher compared to the previous (0.07%
134 - 1.53%) (Supplementary Table 3) sika deer genome assemblies(Han et al. 2022; Xing et al.
135 2023). Utilizing a range of methods, including RNA-seq and Iso-Seq transcriptome data
136 (Supplementary Table 4), gene projection based on genome alignment, protein homology
137 annotation, and *de novo* annotation, we annotated a total of 22,119 protein-coding genes,
138 including 13 mitochondrial genes. The number of genes annotated was comparable to the
139 RefSeq gene annotation of red deer. Functional annotations were obtained for 20,715 of these
140 genes. The BUSCO completeness score of 99.5%, surpasses all previous sika deer
141 annotations(Han et al. 2022; Xing et al. 2023; Qin et al. 2023) and annotations from published
142 Cetartiodactyla reference genomes such as those of cattle, vaquita, and pig(Talenti et al. 2022;
143 Morin et al. 2021; Warr et al. 2020) (Fig. 1C). Moreover, we identified 35 genes in our
144 annotation that spanned more than 1 Mb, a rarity in previous annotations due to the splitting
145 strategy. Finally, by integrating short-read and long-read transcriptomic data, we annotated 5'
146 UTR for 14,979 genes and 3' UTR for 15,760 genes.

147 By utilizing our assembly and annotation to reanalyze previous single-cell data, we
148 found a substantial increase in the number of detected cells and genes compared to our earlier
149 version (Qin et al. 2023) (Fig. S2) (Supplementary Table 5). Furthermore, the integration of
150 single-cell quality control using mitochondrial gene data resulted in a more precise cell atlas
151 (Fig. 1D, Fig. 1E). During our quality control process, fibroblast populations with unknown
152 functions identified in our previous study(Qin et al. 2023) were excluded due to their
153 association with low-quality cells characterized by high mitochondrial gene content and low

154 read counts. We also successfully annotated T cells and mast cells for the first time in the
155 context of antler regeneration. Furthermore, we discovered a proliferative subgroup of PMC
156 (pPMC) that abundantly appeared at dac5 (days after casting), expressing both *TNN* (ABPCs
157 marker) and *ACTA2* (pericytes marker). This suggests that these cells are activated stem cells
158 with the potential for both osteochondrogenesis and angiogenesis. In other words, the pPMCs
159 may represent the intermediate stem cells from the resting cell to the osteochondrogenesis
160 APBCs and angiogenesis pericytes.

161 **Chromatin accessibility landscapes and gene regulation networks for organs of sika deer**

162 To construct regulatory networks involved in antler regeneration, we collected diverse
163 samples to train the PECA2 model. In total, we performed RNA-seq and ATAC-seq
164 sequencing for a total of 32 samples including two replicates for the lung, spleen, liver,
165 adipose tissue, muscle, skin, rumen, and antler pedicle periosteum (PP) at four distinct stages
166 of antler regeneration (Fig. 2A) (Supplementary Table 4,6). The PP samples were the same as
167 those used in previously published single-cell studies on antler regeneration, which
168 encompassed critical stages of antler regeneration. The ATAC-seq samples achieved an
169 average of 47,686,163 unique read pairs, with enrichment observed at transcription start site
170 (TSS) (Fig. S3). Overall, we identified 178,269 non-redundant open chromatin regions
171 (OCRs) in these tissues. Among these, 41,142 OCRs were activated in the PP tissue, with
172 15,398 OCRs activated specifically in PP (Fig. S4A). Compared to all OCRs, PP-specific
173 OCRs were less abundant in promoter regions, with a larger proportion found in distal

174 intergenic regions (Fig. S4B). This suggests that PP-specific OCRs are more likely to function
175 as distal elements during regeneration. When projected onto the red deer genome, 98.1%
176 OCRs were fully mapped (Fig. S4C), which is much higher than the ratio of synteny region
177 (89.1%) among the global genomes between the two species, suggesting these identified
178 OCRs are highly conserved and thus may have functional constraint. The 1,808 OCRs that
179 were absent or fragmented in red deer were considered specific to sika deer. Additionally,
180 4,250 potential target genes of these 1,808 sika deer-specific OCRs identified by Pearson's
181 correlation analysis were enriched in immune-related functions (Fig. S4D). Through
182 hierarchical clustering and PCA analysis, we discovered notable distinctions between the PP
183 and other tissue types, for both the ATAC-seq and RNA-seq datasets (Fig. 2B-C, Fig. S5).
184 Meanwhile, the skin and rumen, both of which have a similar stratified squamous epithelium
185 composition (Pan et al. 2021), clustered together, indicating that our data accurately capture
186 the similarities between these tissues in the sika deer.

187 To identify regulators driving the variance of chromatin accessibility, we conducted TF
188 (transcription factor) motif enrichment analysis on the peaks from each tissue (Fig. 2D). We
189 identified the TF motifs with the most significant enrichment for each organ and found that
190 these TFs played important roles in the corresponding organs of model organisms, such as
191 HNF4A in the liver (Radi et al. 2023) and MEF2s in the muscle (Taylor and Hughes 2017).
192 The regulatory pattern of PP tissues was similar to that of skin and rumen, contrasting with
193 the results of ATAC-seq and RNA-seq analysis. This similarity was mainly attributed to the
194 binding of AP-1s, KLFs and CNCs, which are TF families involved in cell proliferation and

195 stimulus response. Meanwhile, the specific motif enrichment of the RUNX family in PP
196 tissue indicates a strong osteogenic potential of PP tissues, which differs from epithelial
197 tissues.

198 To establish accurate TF-regulatory element (RE) and RE-target gene (TG) regulatory
199 relationships, we used PECA2 to model the regulatory network. These tissue-specific
200 networks contain an average of 6190 TGs and 318 TFs. The TGs of PP-specific REs were
201 enriched in antler-specific genes identified in a previous study(Wang et al. 2019b) (Fisher's
202 exact test, p -value= 1.64×10^{-16}), suggesting our model can identify important REs function in
203 gene co-option for antler regeneration. We performed Kleinberg's HITS analysis(Kleinberg
204 2000) to calculate hub scores which representing significance of TFs in network. We used
205 Enrichr with 'ARCHS4 tissue' database (Xie et al. 2021; Lachmann et al. 2018)to perform
206 enrichment analysis with the top 50 TFs with highest hub scores for each organ. The results
207 for most major organs matched with corresponding human organs (Supplementary Table 7),
208 indicating that our network could accurately identify the important TFs that contribute to
209 tissue specificity. Tissue enrichment of PP tissues in four stages matched with omentum, an
210 organ with strong regenerative potential(Di Nicola 2019) , suggesting that tissues with strong
211 regenerative ability share similar TF regulatory patterns.

212 **Hub TF dynamics in antler regeneration**

213 Hierarchical clustering and TF-binding motif analysis of multi-omics data revealed two
214 main stages of deer antler regeneration (Fig. 2C, D, Fig. 3A). The first stage, encompassing

215 *dac0* and *dac2*, involved stem cell activity or early development (NFYs, SPs) and stimulus
216 response (CREBs, CNCs, ZNF189, MITF, AP-1s, and sMAFs) as the main biological
217 processes. From *dac5* onwards, TF families related to development (RUNXs, NFATs, TEADs,
218 MEF2s) predominated, suggesting that stem cells had activated and initiated developmental
219 processes such as chondrogenesis. These findings align with a previous study on *dac5* PP
220 ectopic ossification and confirm that *dac5* is a critical turning point for antler
221 regeneration (Qin et al. 2023). We then selected the hub TFs (those with the top 10 highest hub
222 scores) of GRNs at each stage to identify the key TFs in regeneration (Fig3 B). Similar to the
223 TF motif patterns, some hub TFs are shared with other tissues, such as *KLF4* in lung and skin
224 (Supplementary Table 8). Notably, *KLF4*, *JDP2*, *MYC*, and *NFE2L2* are hub TFs across all
225 four stages, suggesting they serve as core TFs throughout the entire regeneration process. The
226 hub TFs identified in the early regeneration stages were primarily associated with stimulus
227 response (*RXRA*, *RXRB*, *NR3C1*), while those appearing later were mostly related to specific
228 developmental processes such as chondrogenesis (*RUNX2*, *SP7*), angiogenesis (*SP3*). We also
229 identified neural crest-related TFs (*TFAP2A*, *TEAD2*), which supports the earlier hypothesis
230 that antler originate from the neural crest (Wang et al. 2019b).

231 We observed a high degree of overlap among the regulons of hub TFs across all stages
232 (Fig. S6). Similarly, the expression profiles of hub TFs in the cellular atlas of antler
233 regeneration reveal that most hub TFs are broadly expressed across various cell types, with
234 only a few exceptions (Fig. S7). For example, *SP7* is specifically expressed in activated stem
235 cells and osteoblasts. These findings suggest that the identified hub TFs may function

236 throughout the entire antler regeneration process through combinatorial coordination.

237 To further elucidate the stage dynamics of the antler regeneration process, we conducted
238 functional enrichment analysis on the shared TGs of these hub TFs (Fig. 3C). The PI3K-AKT
239 signaling pathway and collagen expression-related terms such as Extracellular Matrix
240 Organization were enriched across all four periods. During the dac0 period, there was notable
241 enrichment in Mesenchymal Cell Differentiation, indicating an early stem cell response to
242 antler casting. Both the dac0 and dac2 periods exhibited similarities, with enrichment in terms
243 associated with immune and stress responses, including the TNF signaling pathway, Antigen
244 Processing and Presentation of Exogenous Peptide Antigen, and Wound Healing, indicating a
245 response to wound exposure following antler casting. Significantly, both periods also showed
246 enrichment in Osteoclast Differentiation, aligning with previous studies that suggest
247 osteoclasts drive antler casting(Goss et al. 1992). In the dac5 period, when regenerative
248 APBCs appeared, enrichment was observed in developmental functions such as Skeletal
249 System Development and Embryonic Organ Development. Additionally, the emergence of the
250 Relaxin signaling pathway, known for its role in inhibiting inflammation and
251 fibrosis(Valkovic et al. 2019), marked the transition from an immune response to the initiation
252 of the regenerative process during this period. The subsequent dac10 period showed
253 enrichment in collagen and skeleton development-related terms. In summary, our networks
254 provide a detailed explanation of the antler regeneration process, illustrating the development
255 from antler stem cells to different tissues activating at dac5, driven by the immune response to
256 antler casting.

257 As regeneration developmental process activated at *dac5*, we extracted the subnetwork
258 of hub TFs at *dac5* (Fig. 3D). This subnetwork has a clear hierarchical structure and consists
259 of the time course pattern. The hub TFs related to stem cell and stimulus response in the
260 initial stages are in the upstream of the subnetwork, while the hub TFs related to
261 developmental processes are in the downstream of the subnetwork. Among the 4 core factors,
262 *MYC* is located at the top of the subnetwork and *KLF4* has the most regulatory connections in
263 the subnetwork, indicating a crucial role in generation activation of these two Yamanaka stem
264 cell factors. Sika deer antler stem cells have shown higher stemness and proliferation stability
265 (over 20 passages) than regular mesenchymal stem cells(Seo et al. 2014; Binato et al. 2013).
266 Moreover, compared to the mouse digit tip regeneration, deer antler can regenerate many
267 more times(Dolan et al. 2022). By comparing the stem cells from deer antler PP, mouse
268 regenerative digit tip (P3), and non-regenerative digit tip (P2) (Fig. 3E), we found different
269 quiescent stem cells from antler and P3 differentiated into similar activated stem cell to
270 initiate regeneration (Fig. S8A, B). Specifically, *KLF4* shows differential expression between
271 antler and P3 (p -value $< 1 \times 10^{-300}$) by Wilcoxon analysis, while *MYC* is specifically expressed
272 in deer antler (Fig. 3F). Comparative analysis between human long bone and calvaria bone
273 gene expression data(He et al. 2021) shows the different expression is not dominated by the
274 difference between endochondral ossification (P2 and antler) and intramembranous
275 ossification (P3) (Fig. S8C), suggesting that the high expression levels of these two factors
276 may contribute to the strong stemness of deer antler stem cells. Additionally, we examined the
277 evolutionary status of their network-related REs and found a deer-specific open element

278 (Chr14:13194714-13195323) with a deer-specific insertion upstream of *MYC* (Fig. 3G). This
279 element contains a 7 bp deer-specific insertion in a conserved region, and our ATAC-seq data
280 for different organs showed that it was exclusively open in PP tissue. Our dual luciferase
281 reporter experiment demonstrates that following the cervid-specific insertion, this element
282 exhibits a significantly enhancer effect than the wild-type RE with the pGL4.23 minimal
283 promoter as a control (Fig. 3G). These findings suggest that the emergence of this stronger
284 regulatory element has led to the recruitment of *MYC* in the deer antler stem cells,
285 contributing to their strong stemness.

286 **cTOP modeling uncovers regulation of cellular programs in antler regeneration**

287 Although the hub TFs and their regulons have revealed the dynamics of gene regulatory
288 profiles during antler regeneration, the regulation of cellular program dynamics remains
289 unclear. To address this, we coupled our TF combinatorial regulatory modules identified from
290 bulk GRNs with single cell expression profiles to explore the cell subpopulation
291 heterogeneity. We developed a model called cTOP (combinatorial TF Oriented Program) (see
292 details in Methods), which extracts cellular programs including combinatorial TF module and
293 their specific TGs from single-cell RNA-seq data (Fig. 4A).

294 We used the cTOP model to analyze the single-cell data at PP *dac5*, a stage when
295 regeneration has been activated. In total, we identified six Cellular Programs mediated by TF
296 Combinatorial Regulations (CPCRs). Uniform manifold approximation and projection
297 (UMAP) dimensionality reduction revealed that the cell types assigned by the traditional

298 clustering method (Fig. 4B) closely matched with the cellular program assignments based on
299 the highest CPCR score (Fig. 4C), except for CPCR3 which was consistent with proliferating
300 cells across different cell types. The strong correlation between functional enrichment (Fig.
301 4D) and cell classification demonstrated the accuracy and sensitivity of our method in
302 extracting cellular programs. Furthermore, our findings revealed that cell types were not
303 always defined by individual CPCRs, but rather by different compositions of multiple
304 CPCRs. For instance, CPCR2 was detected in nearly all cell types (Fig. S9), indicating the
305 involvement of immune responses triggered by antler casting in various biological processes
306 during antler regeneration. Therefore, in our model, cell types are determined by the
307 composition of CPCRs, and cell differentiation can be described as the dynamic interplay of
308 CPCRs. We also applied the cTOP model to liver data. Our model successfully identified
309 metabolism, immune, and angiogenesis programs in the relevant cell types (Fig. S10 A-E).
310 Notably, the programs associated with endothelial cells in both PP and liver tissues shared 9
311 TFs (*AR*, *ID2*, *GATA2*, *KLF4*, *MAFK*, *MAFG*, and *PPARA*) and 55 TGs tightly related to
312 angiogenesis (Torres-Estay et al. 2015; Sangwung et al. 2017; Lasorella et al. 2005; Dong et
313 al. 2020) (Fig. S10F), suggesting that the cTOP model is robust in identifying similar
314 regulatory programs across different tissues.

315 Deer antler regeneration is a stem cell-based epimorphic process (Wang et al. 2019a).
316 Within the stromal cell lineage which includes PMCs, pericytes, APBCs, and osteoblasts, we
317 identified five CPCR cell programs: CPCR1 (osteochondrogenesis), CPCR2 (stimulus
318 response), CPCR3 (cell proliferation), CPCR4 (mesenchymal stem cell phenotype), and

319 CPCR5 (pericyte-related angiogenesis). Pseudotime trajectories demonstrate that
320 mesenchymal stem cells differentiate into pericytes and APBCs through proliferative PMCs
321 (Fig. 4E, Fig. S11 A-D). Throughout this process, the activity of CPCR4 gradually decreases,
322 while CPCR3 is activated. Finally, the dominant program transitions into CPCR1 and CPCR5,
323 respectively (Fig. 4F, Fig. S11 E-F).

324 At the beginning of differentiation, we found that CPCR4 functioned more as a PMC
325 activation program rather than a PMC resting program, as inferred from the cell annotation.
326 The TFs in this subnetwork include TF families (NFATCs and TEADs) that emerge at *dac5* by
327 motif enrichment analysis and regulate the osteochondrogenic TFs (*RUNX2*, *SOX9* and *EN1*)
328 of the dominant program in ABPCs: CPCR1 (Fig. 4G). Core TFs in the CPCR4, including
329 *FOSL1*, *FOS*, *PRDMI*, *BACH1* and *NFATC1* express higher at the first two days after casting,
330 suggesting their early roles in regulating stem cell activation (Fig. 4H). TFs in the AP-1 family
331 including *FOS* and *FOSL1* have been found activating stem cells in muscle regeneration, and
332 NFACT1 is crucial for ensuring bone repair and regeneration in skeletal stem cells (Yu et al.
333 2022) . These TFs may be key factors for stem cell activation by regulating genes in WNT
334 and Hedgehog signaling pathway.

335 We further found regulatory identified between programs following stem cell activation
336 (Fig. 4J). The proliferation program CPCR3 was regulated by *KLF4*, *ID2/E2Fs*, *CREBs*, and
337 *SMADs* as well as *SP7* from CPCR1 and *SIX2* from CPCR5. Expression of *SIX2* and *SP7*
338 specifically increased along differentiation of angiogenesis and osteochondrogenesis,
339 respectively (Fig. 4I). *SP7* is a well-known master regulator for skeleton development while

340 *SIX2* is important for cranial skeleton development from cranial neural crest stem cell(Liu et
341 al. 2019b). Additionally, *CPCR5*, the pericyte-dominant program, also regulated the *CPCR1*,
342 suggesting multifunctional roles for pericytes during antler regeneration. These findings
343 indicate that TFs expressed by the differentiation programs can, in turn, regulate the stem cell
344 proliferation program, and that *SIX2* and *SP7* may determine the fate of stem cell
345 differentiation through this regulatory mechanism.

346 We also found significant effects of TGF- β pathway on different cellular programs
347 involved in antler regeneration. In a mouse heterotopic ossification model, high levels of
348 *TGFB1* induces the fibroblasts differentiating to chondrogenic progenitor cells(Sorkin et al.
349 2020). We observed a similar role of *TGFB1* in antler osteochondrogenesis in contrast to
350 *TGFB2/3*. The expression of *TGFB2* and *TGFB3* peaks at the onset of osteochondrogenic
351 differentiation and then decline, while the expression of *TGFB1* increases during this process
352 (Fig. 4F). *TGFB2* and *TGFB3* are crucial for the maintenance of the valvular interstitial cell
353 phenotype, a multipotent cell type in heart valves(Wang et al. 2021). Additionally, *TGFB2* has
354 been reported to inhibit osteogenesis of mesenchymal stem cells, contrasting with the role of
355 *TGFB1*(Li et al. 2022). These results suggest a critical role of *TGFB2/3* in PMCs stemness
356 maintenance by suppressing differentiation. Similar interplays were also found between TGF-
357 β pathway related TFs like *ID2* and *SMADs* in *CPCR3*. In airway basal stem cells, the TGF- β -
358 *ID2* axis was reutilized to promote tissue regeneration, with overexpression of *ID2* leading to
359 a tumorigenic phenotype(Kiyokawa et al. 2021). *SMADs* are canonical downstream factors of
360 TGF- β pathway. *SMAD4*, in combination with TGF- β activated *SMAD2* regulates DNA

361 repairing and cell cycle to inhibit tumorigenesis. We found these factors regulating
362 proliferative genes like *UBE2C* and *TOP2A* together. The coordination between *ID2/E2Fs*
363 and SMADs may keep the balance between regenerative ability and cancer suppression in
364 antler generation.

365 In summary, our cTOP approach effectively modeled the cellular dynamics of the stem
366 cell differentiation process in antler regeneration. By constructing the subnetwork of CPCRs,
367 we have found key factors operating within specific cellular programs and the interactions
368 between these programs.

369 **Discussion**

370 The multiomic approach has become an efficient method for resolving complex
371 biological processes. However, the application of such methods in non-model species has
372 been hindered by the lack of high-quality omics data. This study generated the high-quality
373 and comprehensive genomic data of sika deer, making it suitable for subsequent multiomic
374 studies. We observed that the genome size estimated for flow cytometry is larger than those
375 from *k*-mer estimation as well as previous assembly based on technologies like ONT and
376 high-throughput sequencing, a discrepancy also noted in other ruminant species such as cows
377 and reindeer(Kent et al. 1988; Goss et al. 1992). Our analysis indicates that this gap is mainly
378 due to the failure of anchoring massive centromeric sequences, a characteristic feature of
379 ruminant genomes. Although we have nearly completed assembling all the euchromatic

380 regions of the sika deer genome, overcoming the large centromeric regions and achieving a
381 complete telomere-to-telomere genome assembly for ruminant species remains a significant
382 challenge. Fortunately, the large gaps in the ruminant centromeric regions contain few or no
383 genes because almost all mammalian genes have been annotated for the sika deer, and
384 therefore they have limited impact on the downstream functional omics analyses.

385 Based on the high-quality sika deer genome, we were able to integrate large amounts of
386 omics data including bulk RNA-seq, single cell RNA-seq, and ATAC-seq data of sika deer to
387 reconstruct the tissue-specific gene regulatory networks and resolve cellular dynamics during
388 antler regeneration. The results reveal that antler PP shared some key regulator, such as AP-1
389 and *KLF4*, with epithelial tissues. AP-1 complexes are prevalent in the regenerative elements
390 of various species, including fruit flies, acoel worms, and bony fish, and are crucial for
391 activating regenerative response enhancers in bony fish(Wang et al. 2020). *KLF4* is essential
392 for the homeostasis and self-renew of epithelial cells in various tissues(Angel et al. 2001;
393 Segre et al. 1999). The recruitment and combination of these TFs may underlie the
394 regenerative potential of antler PP. Several hub TFs, including *MYC*, *KLF4*, *NFE2L2*, and
395 *JDP2*, coordinated the regeneration process from wound healing towards skeletal
396 development. The uncovered cellular regulation dynamics indicated that *FOSL1*, *PRDMI*,
397 and *NFATC1* might drive the stem cell activation and balancing between oncogenic factors
398 *ID2/E2Fs* and anticancer factors *SMADs* in the TGF- β pathway might contribute to the well-
399 organized stem cell proliferation during antler regeneration (Fig. 5e). The proliferation
400 program might also receive potential feedback regulation by *SP7* and *SIX2* from

401 differentiation programs. These findings for the first time revealed the gene and cell
402 regulatory mechanism of deer antler regeneration.

403 Antler progenitor stem cells have been characterized as *PRRX1*⁺ mesenchymal stem cells
404 with certain embryonic stem cell characteristics. The involvement of embryonic stem cell-
405 related TFs like *KLF4* and *MYC* in antler regeneration has been a subject of debate. Our study
406 suggested the hub role of *KLF4* and *MYC* in the regulatory networks of antler regeneration,
407 which could not be easily identified through gene expression analysis alone. Cross-species
408 comparisons highlight their high expression might contribute to the high regenerative capacity
409 of antler stem cells. Along with the identification of cervid-specific regulatory elements and
410 the discovery of stem cell phenotype maintenance by *TGFB2/3*, we have preliminarily
411 elucidated the evolutionary and molecular mechanisms underlying stemness in deer antler
412 stem cells, which may provide valuable insights for regenerative medicine studies.

413 In addition, we proposed the new method, cTOP, in this study to combine our scRNA-
414 seq and bulk ATAC-seq and decode cellular GRNs. Existing gene regulatory network
415 inference methods are not ideally suited for sika deer data. Most methods, such as SCENIC+,
416 GLUE, scReg(Bravo González-Blas et al. 2023; Duren et al. 2022; Cao and Gao 2022)
417 require paired scRNA-seq and scATAC-seq, which is expensive and very sensitive to quality
418 of library preparation. Other methods, such as SCENIC and GRNBoost(Aibar et al. 2017; P
419 assemiers et al. 2022) construct a coarse network using only scRNA-seq. Compared to cTOP,
420 scRNA-seq-only methods, which infer regulatory relationships primarily from co-expression
421 at the single-cell level, may identify many false-positive regulations among TFs and TGs

422 within a co-expression module. The approach of cTOP, decomposing TF modules from bulk
423 GRNs and then coupling with single cell expression profile, offers a cost-effective alternative
424 to obtain highly confident cellular GRNs. Importantly, cTOP fully utilized all data generated
425 in a non-model animal. In future, we will generalize cTOP method to scRNA-seq data only
426 cases and compare with existing methods. Furthermore, cTOP can easily be used in model
427 animal-based development and disease research with even better performance, given the fact
428 that the original PECA2 method was developed with larger and more diverse paired
429 expression and chromatin accessibility data from mouse and human.

430 In conclusion, we have identified the key factors and pathways in stem cell activation,
431 proliferation and differentiation of antler regeneration through cellular GRN modeling based
432 on our high-quality omics data. The gene regulatory mechanism underlying the strong
433 regenerative capacity and delicate balance between regeneration and cancer suppression in
434 cervids will provide new clues for both regeneration and cancer medicine studies.

435

436 **Methods**

437 **Sample collection and genome sequencing.**

438 Four 2-year-old male sika deers (*Cervus nippon*) were used for sampling regenerating
439 antler Tissues, on days 0, 2, 5 and 10 after casting. Another 2-year-old male sika deer was
440 sacrificed for sampling normal organs. Blood from 4 chickens, sika deers and rats was

441 collected to conduct flow cytometry for genome size estimation (detail in Supplemental
442 Methods).

443 Genomic DNA was extracted from liver using the standard cetyltrimethylammonium
444 bromide method. For HiFi sequencing, SMRTbell library construction and sequencing were
445 performed at Novogene (Tianjin, China) or BerryGenomics (Beijing, China) following the
446 official protocols of PacBio for preparing ~20-544kb SMRTbell libraries. For Hi-C
447 sequencing, we followed the standard protocol described previously with minor modification,
448 using the same sample with HiFi sequencing(Belton et al. 2012) at Novogene (Tianjin,
449 China). ATAC-seq was performed by standard protocol as previously reported(Buenrostro et
450 al. 2013) at Novogene (Tianjin, China). Details of DNA and RNA library preparation are
451 described in Supplementary Methods, and statistics of all data collected for each bat are
452 provided in Supplementary Table 1 and 3.

453 **Genome assembly.**

454 Hifiasm (v0.19.5) was used to assemble the HiFi reads with ONT reads we generated
455 previously, and Hi-C reads from same individual to generate phased contigs. Then the contigs
456 of each haplotype were merged and scaffolded with Juicer(Durand et al. 2016) and
457 3DDNA(Dudchenko et al. 2017) to check switch error of phasing (most in sex chromosomes)
458 (Fig.S1 D). Then each haplotig was scaffolded separately to increase scaffolding accuracy
459 with higher Hi-C contact resolution.

460 Mitochondrial genome was assembled and annotated using MitoHiFi (v3.2.1)(Uliano-

461 Silva et al. 2023) with HiFi reads.

462 We assessed genome completeness and consensus quality value (QV) using
463 Merqury(Rhie et al. 2020). Besides, we performed a BUSCO(Waterhouse et al. 2018)
464 assessment of the genome sequences using the certa_odb10 database.

465 **Genome annotation.**

466 RepeatMasker(Tarailo-Graovac and Chen 2009) was first used to detect and mask the
467 repetitive region. Then we integrated different evidence to predict genes. First, we use
468 GeMoMa(v1.9) to do homology-based annotation with annotation of human, cattle, red deer
469 and yarkand deer from NCBI or Ensembl as reference. Second, we processed whole genome
470 alignment to these related species genomes with UCSC chain/net pipeline(Kent et al. 2003)
471 and projected gene annotations using TOGA (v1.1.7)(Kirilenko et al. 2023). Third, we used
472 transcriptome workflow in REAT (v0.6.1 <https://github.com/EI-CoreBioinformatics/reat>) to
473 integrate short and long transcriptomic data and generate a highly confident gene model. De
474 novo prediction was applied with BRAKER (v2.1.5)(Hoff et al. 2016) in with transcriptomic
475 model as hint. All evidence was integrated using MAKER (v3.01.03)(Campbell et al. 2014).

476 **ATAC-seq data process.**

477 ATAC-seq reads were cleaned by fastp (v0.23.1)(Chen et al. 2018) and then were
478 aligned to the reference genome using Bowtie 2(Langmead and Salzberg 2012). These reads
479 were then filtered for high quality (MAPQ \geq 13), we also removed reads that were not

480 properly paired and with PCR duplicates by Picard (version 2.25.7
481 <https://broadinstitute.github.io/picard/>). All peak calling was performed with MACS2 (version
482 2.1.0)(Zhang et al. 2008) using “–call-summits nomodel –shift -100 –extsize 200”. Motif
483 enrichment was performed by HOMER(Heinz et al. 2010).

484 **RNA-seq data process.**

485 Short RNA-seq reads were cleaned by fastp and then were aligned to the reference
486 genome with HISAT2(Kim et al. 2019). The mapped reads of each sample were assembled by
487 StringTie (v1.3.3b)(Kovaka et al. 2019).

488 Iso-Seq reads were preprocessed by ISOSEQ3 (v3.8.0
489 <https://github.com/PacificBiosciences/IsoSeq>). The FLNC reads from ISOSEQ3 were mapped
490 to genome with minimap2 and processed with TAMA pipeline.

491 **Single-cell RNA data process.**

492 10x single-cell RNA-seq data were obtained from our previous research(Qin et al. 2023).
493 Cell Ranger (v6.1.1) was used to make prepare reference and count the gene expression
494 profile. Then we used SCANPY (v1.1.10)(Wolf et al. 2018) to process the analysis. First, we
495 use Scrublet (v0.2.3)(Wolock et al. 2019) to mark and remove doublets. Then we further
496 filtered out cells less than 500 counts or have more than 15% mitochondrial gene expression.
497 After preprocessing, we used harmony (v0.0.9) to integrate data of all stages. We used the
498 Leiden method to cluster cells and the Wilcoxon method to identify marker genes.

499 Pseudotime trajectory inference of *dac5* data was applied with Monocle3 (v.0.5.0)(Setty
500 et al. 2019) and Palantir (v1.3.0)(Setty et al. 2019) separately.

501 **Optimization model to identify TF combinatorial programs by cTOP model.**

502 cTOP is a method to infer TF combinatorial program from scRNA-seq data. In brief, a
503 guidance TF-TG regulatory network is first constructed for a given context. Here we use
504 PECA2 model from paired bulk gene expression and chromatin accessibility to construct TF-
505 TG regulatory network. Second, a guidance TF combinatorial network is built from TF-TG
506 network by connection specificity index (CSI) to represent the similarity of regulons between
507 TFs. Then, we apply non-negative matrix factorization (NMF) based optimization model to
508 both the CSI matrix and the single-cell expression matrix for identifying TF combinatorial
509 program and cell embeddings on TF programs. Three essential steps of cTOP are detailed
510 below.

511

512 **1. Constructing TF-TG regulatory network and estimating TF combinatorial** 513 **effect**

514 We use the PECA2¹² model to build a TF-TG regulatory network. PECA2 takes paired
515 gene expression (bulk RNA-seq) and chromatin accessibility data (bulk ATAC-seq) as input,
516 with replicant merged for each tissue. The prior data of PECA2, including TF-TG correlation
517 and RE-TG interaction, is calculated from data of multiple deer organs (Supplementary Table
518 S5) for deer specific regulatory network. The output of PECA2 is a TF-TG regulatory

519 strength matrix, denoted by R matrix, for M TFs and N TGs. Specifically, R_{ij} , which is
 520 the i -th row and j -th column of TRS matrix, is the regulatory strength score of the i -th TF
 521 on j -th TG.

522 Then we use the connection specificity index (CSI)(Bass et al. 2013) to assess the
 523 combinatorial effect of two TFs. For the i -th and j -th TFs, we use R_i and R_j to represent
 524 their regulatory strength scores across all the TGs. Then we calculate the Pearson correlation
 525 of their regulatory strength $PCC_{ij} = PCC(R_i, R_j)$. Then CSI score considers the specificity of
 526 Pearson correlation to evaluate the combinatorial effect of i -th and j -th TFs:

$$527 \quad CSI_{ij} = \frac{\#\{l: PCC_{il} \leq PCC_{ij} - \varepsilon, PCC_{jl} \leq PCC_{ij} - \varepsilon\}}{M} \quad (1)$$

528 Here M was the total number of TFs. ε was a constant with a default value of 0.05. A
 529 high CSI score indicated that two TFs specifically regulated a group of TGs.

530 2. Identifying TF combinatorial program from scRNA-seq data

531 We use cTOP model to identify TF combinatorial program from single cell RNA-seq
 532 data. There are three inputs of the cTOP model: 1) TF combinatorial network represented by
 533 above CSI matrix, 2) TF-TG regulatory network represented by above trans-regulatory score
 534 matrix R , and 3) single cell gene expression matrix E . We expected the gene expression
 535 matrix E to be the coupled with TF combinatorial programs through regulatory network R .
 536 Formally, the cTOP optimization model is formulated as follows.

$$537 \quad \min_{X, H} \|C - XX^T\|_F^2 - \mu_1 C \circ (XX^T) + \mu_2 \|E - WH\|_F^2 - \mu_3 \text{tr}(X^T R W)$$

$$538 \quad s. t. X \geq 0, H \geq 0, \sum_i x_{ik}^2 = 1, k = 1, 2, \dots, K; \sum_k x_{ik} = 1, i = 1, 2, \dots, M \quad (1.)$$

539 The cTop model has three components:

540 1) $\|C - XX^T\|_F^2 - \mu_1 C \circ (XX^T)$: The first two terms are designed to detect TF
 541 modules from TF-TF combinatorial effect matrix. The first term decomposes TF-TF
 542 combinatorial matrix for detecting TF combinatorial modules and the second term
 543 constrains the detected TF modules to be TF combinations with large CSI scores. This
 544 component will output variable X , which is a M by K matrix to reveal the combinatorial
 545 effect of M TFs in K TF combinatorial programs. We used X to obtain TF modules of
 546 TF programs. Given the TFs' combinatorial effect $X_k = (X_{1k}, X_{2k}, \dots, X_{Mk})^T$ in the k -th
 547 combinatorial regulon, we computed the combinatorial effect of i -th TF and j -th TF in
 548 k -th TF programs:

$$549 \quad CE_{ij}^k = X_{ik} \times X_{jk} \quad (2.)$$

550 We assumed the combinatorial effect of TF pairs in each TF programs followed Gamma
 551 distribution. We used threshold $P\text{-value} \leq 0.01$ to select TF pairs for k -th TF programs and
 552 the significant TF pairs formed the TF module of k -th TF programs.

553 2) $\|E - WH\|_F^2$: the third term was to cluster and obtained gene expression
 554 programs for scRNA-seq. This component will output matrix W and H . W was a N
 555 by K matrix to represent the gene expression program and each column of W indicated
 556 the mean expression of TGs regulated by the corresponding TF module. And W was used
 557 to obtain the TGs of each TF program by gene expression. H was a K by c matrix to
 558 reveal assignment weights of c cells for K TF programs. We assigned each cell to a TF
 559 program with the largest assignment weight.

560 3) $\text{tr}(X^T RW)$: The last two terms exerted specificity on TF program by coupling

561 the TF modules with the gene expression programs through the regulatory network. This
 562 component gave constraints to TF modules: the TF modules should regulate TGs that have
 563 specific expression in certain cell types/states. This constraint enabled TF programs to
 564 utilize not only the specificity of TFs but also the specificity of Res-TGs and TF
 565 combinatorial effect to identify core TF combinations for cell type/states.

566 These three outputs (X , W , H) would be used for describing TF combinatorial
 567 programs. We modeled the k -th TF combinatorial program as follows.

$$568 \quad (X_k(W_{1k} + W_{2k})^T) \circ R \quad (3.)$$

569 Here X_k and W_k were the k -th column of X and W , respectively. The k -th TF
 570 combinatorial program was represented by the TF module defined by X_k with equation (6).
 571 The TGs of k -th TF combinatorial program were given by W . The k -th cell state, which
 572 was cells governed by k -th TF combinatorial program, was defined by H .

573 **3. Annotating cell clusters with linear combinations of TF combinatorial** 574 **programs**

575 A cell cluster is a group of cells in scRNA-seq data. For a cell cluster, we suppose this
 576 cell cluster has n cells $G = \{g_1, g_2, \dots, g_n\}$. We represent this cell cluster as the TF
 577 combinatorial program with the averaged coefficients of all the cells in G , respectively:

$$578 \quad D = \frac{1}{n} \sum_{i=1}^n H_{g_i} \quad (4.)$$

579 where H_{g_i} is the columns corresponding to the cell g_i . Then the TF combinatorial
 580 program combination coefficients of the given cell cluster will be D .

581 **4. Model initiation, parameter selection, and optimization algorithm**

582 To initiate our model, we first solved the component (1) and (2) of our model
583 independently. These three components gave us the initiation of five variable: X^0 , W^0 , H^0 .

584 The initiation matrix of three variables enabled us to determine the hyper-parameters in
585 our model. There were three hyper-parameters in our model: μ_1 , μ_2 , μ_3 , and K . μ_1 , μ_2 ,
586 μ_3 , were parameters to balance the scale of five terms in our model, which could be
587 determined by the initiation matrix:

588
$$\mu_1 = \|C - X^0 \cdot X^{0T}\|_F^2 / (C \circ (X^0 \cdot X^{0T})) \quad (5.)$$

589
$$\mu_2 = \|C - X^0 \cdot X^{0T}\|_F^2 / \|E - W^0 \cdot H^0\|_F^2 \quad (6.)$$

590
$$\mu_3 = \|C - X^0 \cdot X^{0T}\|_F^2 / \text{tr}(X^{0T} \cdot R \cdot W^0) \quad (7.)$$

591 The hyper-parameter K was the number of TF combinatorial programs, which was
592 consistent with number of TF modules in C and number of cell type/states for single cell
593 data. K could be determined in two ways. First, if we had prior knowledge about the number
594 of TF modules or cell types, we could direct assign this number to K . Second, if we don't
595 have biological insights about the data beforehand, we could try different K to find modules
596 and compute modularity in the TF combinatorial effect matrix to select the best one. And the
597 number of K could also be determined by a method similar to that in Brunet et al.

598 Starting from the initiation matrices and hyper-parameters, we proposed a multiplicative
599 update algorithm to solve the optimization problem of the cTOP model. We used X_{ij} to
600 represent the element of the i -th row and the j -th column in matrix X and W_{ij} and H_{ij} be
601 the corresponding elements in W and H . We adopted the following update roles and stopped

602 the iteration when the relative error is less than 0.0001.

$$603 \quad X_{ij} = X_{ij} \cdot \frac{(4 + 2\mu_1)C \cdot X + \mu_4 R \cdot W}{4X \cdot X^T \cdot X} \quad (8.)$$

$$604 \quad W_{ij} = W_{ij} \cdot \frac{E \cdot H^T + \mu_3/\mu_2 R^T \cdot X}{W \cdot H \cdot H^T} \quad (9.)$$

$$605 \quad H_{ij} = H_{ij} \cdot \frac{W^T E}{H^T W H} \quad (10.)$$

606

607 **Cervid specific structure variation identification.**

608 To check the structure variants in elements, we first used UCSC chain/net pipeline to
 609 generate whole genome alignment to rein deer (HranTar1). Then we used liftOver to map
 610 elements from sika deer to rein deer. Then we extract maf file with rein deer as reference from
 611 the hal file generated by cactus in Zoonomia project(Genereux et al. 2020). An in-house script
 612 called SV_caller was used to find structure variants follow these criterions: 1. Identity of
 613 sequence for ingroup species in the SVs is not below 90%. 2. Identity of sequence for all
 614 species 50bp flank SVs is not below 50%. 3. Length of SV is over 5bp.

615 **Comparative analysis of OCRs.**

616 To check the evolutionary conservation of OCRs, we first used UCSC chain/net pipeline
 617 to generate whole genome alignment to red deer (mCerEla1.1). Then we used liftOver to map
 618 elements from sika deer to red deer. Potential target genes were identified with Pearson's
 619 correlation test. Functional enrichment was conducted with Enrichr.

620 **Luciferase reporter assay.**

621 The dual luciferase reporter constructs engineered in this study were developed on the
622 pGL4.23[luc2/minP] Vector (E8411, Promega), pGL4.74[hRluc/TK] Vector(E6911,
623 Promega) which were served as the base plasmid. The Dual Glo Luciferase luminescent assay
624 (E1960, Promega) was carried out in accordance with the manufacturer's protocol with
625 slight modifications. Detailed protocols for transient transfection and stable measurements are
626 described here. After transient transfection in HEK-293T adherent cells, discard the cell
627 culture medium and rinse the cells twice with PBS. Add 120ul 1×passive lysis buffer (E194A,
628 Promega) to each well to lyse the cells. After incubation at room temperature for 10 min,
629 lysates were transferred to 96-well flat-bottomed white polystyrene plates (3912, Corning).
630 Using the automatic loading function of a multifunctional microplate reader (Synergy Neo2,
631 BioTek), add 50ul of firefly luminescence detection solution and 50ul of sea cucumber
632 luminescence detection solution according to 2.5ul of the sample to be tested in each well,
633 and load the sample and detect the final luminescence value.

634 **Data access**

635 All raw sequencing data generated in this study have been submitted to the Genome
636 Sequence Archive of China National Center for Bioinformation (GSA;
637 <https://ngdc.cncb.ac.cn/gsa>) under accession number CRA018294 (single cell RNA-seq data),
638 CRA018238 (genomic data), and CRA015420 (RNA-seq and ATAC-seq data). Genome

639 assembly was submitted to NCBI GenBank database with accession number
640 GCA_038088365.1. Genome assembly, gene annotation and network files were also updated
641 to Zenodo database with <https://zenodo.org/records/13298156>. The Source code is provided with
642 this paper in Supplementary Codes. The source codes and sample data for cTOP are available
643 at <https://github.com/AMSSwanglab/cTOP>. The codes for structure variation are available at
644 https://github.com/lizihe21/SV_caller.

645 **Competing interest statements**

646 The authors declare no competing interests.

647 **Acknowledgements**

648 This work was supported by two grants of NSFC (Nos. 32030016 and 32220103005)
649 and the New Cornerstone Investigator Program to WW. National Key Research and
650 Development Program of China (2022YFA1004800), CAS Project for Young Scientists in
651 Basic Research, and the National Natural Science Foundation of China (12025107) to YW.

652 **Author contributions**

653 W.W., Y.W., Q.Q. and D.J.H conceived and designed the project. Z.H.L., Z.Y.F., W.W.,
654 and Y.W. drafted the manuscripts. Z.H.L., Q.T., B.T.Z., W.W., S.J.J. and L.Z. revised the
655 manuscript. T.Q., G.K.Z. and C.Y.L. collected and prepared the samples, help in assistance

656 with the experiment. J.R.M., G.Q. and H.S.Y. designed and performed all experiments. Z.H.L.
657 Z.Y.X. Z.Y.F. L.Z. and G.H. performed the data analysis.

658 **References**

- 659 Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G,
660 Rambow F, Marine JC, Geurts P, Aerts J, et al. 2017. SCENIC: single-cell regulatory
661 network inference and clustering. *Nature Methods* 2017 14:11 14: 1083–1086.
- 662 Angel P, Szabowski A, Schorpp-Kistner M. 2001. Function and regulation of AP-1
663 subunits in skin physiology and pathology. *Oncogene* 2001 20:19 20: 2413–2423.
- 664 Ba H, Wang X, Wang D, Ren J, Wang Z, Sun HX, Hu P, Zhang G, Wang S, Ma C, et al.
665 2022. Single-cell transcriptome reveals core cell populations and androgen-RXFP2 axis
666 involved in deer antler full regeneration. *Cell Regeneration* 11.
- 667 Bass JIF, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJM. 2013. Using networks
668 to measure similarity between genes: association index selection. *Nat Methods* 10:
669 1169–1176.
- 670 Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a
671 comprehensive technique to capture the conformation of genomes. *Methods* 58: 268–
672 276.
- 673 Binato R, de Souza Fernandez T, Lazzarotto-Silva C, Du Rocher B, Mencialha A,
674 Pizzatti L, Bouzas LF, Abdelhay E. 2013. Stability of human mesenchymal stem cells

- 675 during in vitro culture: considerations for cell therapy. *Cell Prolif* **46**: 10.
- 676 Blaxter ML. 2022. Sequence locally, think globally: The Darwin Tree of Life Project.
677 *Proc Natl Acad Sci U S A* **119**.
- 678 Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I,
679 Christiaens V, Poovathingal S, Wouters J, Aibar S, Aerts S. 2023. SCENIC+: single-cell
680 multiomic inference of enhancers and gene regulatory networks. *Nature Methods* 2023
681 20:9 **20**: 1355–1367.
- 682 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of
683 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-
684 binding proteins and nucleosome position. *Nature Methods* 2013 10:12 **10**: 1213–1218.
- 685 Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome Annotation and Curation
686 Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**: 4.11.1-4.11.39.
- 687 Cao ZJ, Gao G. 2022. Multi-omics single-cell data integration and regulatory inference
688 with graph-linked embedding. *Nat Biotechnol* **40**: 1458–1466.
- 689 Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W, et al.
690 2019. Large-scale ruminant genome sequencing provides insights into their evolution
691 and distinct traits. *Science* **364**.
- 692 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
693 *Bioinformatics* **34**: i884–i890.
- 694 Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. 2022.
695 Haplotype-resolved assembly of diploid genomes without parental data. *Nature*

- 696 *Biotechnology* 2022 40:9 **40**: 1332–1335.
- 697 Chisanga D, Liao Y, Shi W. 2022. Impact of gene annotation choice on the
698 quantification of RNA-seq data. *BMC Bioinformatics* **23**: 1–21.
- 699 Dąbrowska N, Kielbowicz Z, Nowacki W, Bajzert J, Reichert P, Bieżyński J, Zebrowski
700 J, Haczekiewicz K, Cegielski M. 2016. Antlerogenic stem cells: molecular features and
701 potential in rabbit bone regeneration. *Connect Tissue Res* **57**: 539–554.
- 702 Di Nicola V. 2019. Omentum a powerful biological source in regenerative surgery.
703 *Regen Ther* **11**: 182.
- 704 Dolan CP, Yang TJ, Zimmer K, Imholt F, Qureshi O, Falck A, Gregory J, Mayes M,
705 Ritchie K, Koester H, et al. 2022. Epimorphic regeneration of the mouse digit tip is
706 finite. *Stem Cell Res Ther* **13**: 1–12.
- 707 Dong Z, Haines S, Coates D. 2020. Proteomic Profiling of Stem Cell Tissues during
708 Regeneration of Deer Antler: A Model of Mammalian Organ Regeneration. *J Proteome*
709 *Res* **19**: 1760–1775.
- 710 Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS,
711 Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti*
712 genome using Hi-C yields chromosome-length scaffolds. *Science (1979)* **356**: 92–95.
- 713 Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL.
714 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C
715 Experiments. *Cell Syst* **3**: 95–98.
- 716 Duren Z, Chang F, Naqing F, Xin J, Liu Q, Wong WH. 2022. Regulatory analysis of

717 single cell multiome gene expression and chromatin accessibility data with scREG.
718 *Genome Biol* **23**.

719 Duren Z, Chen X, Jiang R, Wang Y, Wong WH. 2017. Modeling gene regulation from
720 paired expression and chromatin accessibility data. *Proc Natl Acad Sci U S A* **114**:
721 E4914–E4923.

722 Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, Juan D,
723 Bejerano G, Casewell NR, Chemnick LG, et al. 2020. A comparative genomics
724 multitool for scientific discovery and conservation. *Nature* **2020 587:7833** **587**: 240–
725 245.

726 Goss RJ, Van Praagh A, Brewer P. 1992. The mechanism of antler casting in the fallow
727 deer. *Journal of Experimental Zoology* **264**: 429–436.

728 Han R, Han L, Zhao X, Wang Q, Xia Y, Li H. 2022. Haplotype-resolved Genome of
729 Sika Deer Reveals Allele-specific Gene Expression and Chromosome Evolution.
730 *Genomics Proteomics Bioinformatics*.

731 He J, Yan J, Wang J, Zhao L, Xin Q, Zeng Y, Sun Y, Zhang H, Bai Z, Li Z, et al. 2021.
732 Dissecting human embryonic skeletal stem cell ontogeny by single-cell transcriptomic
733 and functional analyses. *Cell Res* **31**: 742–757.

734 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh
735 H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors
736 Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol*
737 *Cell* **38**: 576–589.

- 738 Hoang T, Wang J, Boyd P, Wang F, Santiago C, Jiang L, Yoo S, Lahne M, Todd LJ, Jia
739 M, et al. 2020. Gene regulatory networks controlling vertebrate retinal regeneration.
740 *Science* **370**.
- 741 Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1:
742 Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and
743 AUGUSTUS. *Bioinformatics* **32**: 767–769.
- 744 Jimenez E, Slevin CC, Song W, Chen Z, Frederickson SC, Gildea D, Wu W, Elkahloun
745 AG, Ovcharenko I, Burgess SM. 2022. A regulatory network of Sox and Six
746 transcription factors initiate a cell fate transformation during hearing regeneration in
747 adult zebrafish. *Cell Genomics* **2**: 100170.
- 748 Johnston H, Warner JF, Amiel AR, Nedoncelle K, Carvalho JE, Röttinger E. 2021.
749 Whole body regeneration deploys a rewired embryonic gene regulatory network logic.
750 *bioRxiv* 658930.
- 751 Kent M, Chandler R, Wachtel S. 1988. DNA analysis by flow cytometry. *Cytogenet Cell*
752 *Genet* **47**: 88–89. <https://pubmed.ncbi.nlm.nih.gov/3356174/> (Accessed February 27,
753 2024).
- 754 Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution’s cauldron:
755 Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl*
756 *Acad Sci U S A* **100**: 11484–11489.
757 <https://www.pnas.org/doi/abs/10.1073/pnas.1932072100> (Accessed February 20, 2024).
- 758 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome

759 alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*
760 2019 37:8 37: 907–915.

761 Kirilenko BM, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales AE,
762 Ahmed AW, Kontopoulos DG, Hilgers L, et al. 2023. Integrating gene annotation with
763 orthology inference at scale. *Science (1979)* 380.

764 Kiyokawa H, Yamaoka A, Matsuoka C, Tokuhara T, Abe T, Morimoto M. 2021. Airway
765 basal stem cells reutilize the embryonic proliferation regulator, Tgfβ-Id2 axis, for tissue
766 regeneration. *Dev Cell* 56: 1917-1929.e9.

767 Kleinberg JM. 2000. *Hubs, Authorities, and Communities*.

768 Kovaka S, Zimin A V., Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019.
769 Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome*
770 *Biol* 20: 1–13.

771 Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC,
772 Ma’ayan A. 2018. Massive mining of publicly available RNA-seq data from human and
773 mouse. *Nature Communications* 2018 9:1 9: 1–10.

774 Landete-Castillejos T, Kierdorf H, Gomez S, Luna S, García AJ, Cappelli J, Pérez-
775 Serrano M, Pérez-Barbería J, Gallego L, Kierdorf U. 2019. Antlers - Evolution,
776 development, structure, composition, and biomechanics of an outstanding type of bone.
777 *Bone* 128: 115046.

778 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat*
779 *Methods* 9: 357–359.

- 780 Lasorella A, Rothschild G, Yokota Y, Russell RG, Iavarone A. 2005. Id2 Mediates
781 Tumor Initiation, Proliferation, and Angiogenesis in Rb Mutant Mice. *Mol Cell Biol* **25**:
782 3563.
- 783 Li C, Zhao H, Liu Z, McMahon C. 2014. Deer antler – A novel model for studying
784 organ regeneration in mammals. *Int J Biochem Cell Biol* **56**: 111–122.
- 785 Li J, Ge L, Zhao Y, Zhai Y, Rao N, Yuan X, Yang J, Li J, Yu S. 2022. TGF- β 2 and TGF-
786 β 1 differentially regulate the odontogenic and osteogenic differentiation of
787 mesenchymal stem cells. *Arch Oral Biol* **135**.
- 788 Liu R, Low WY, Tearle R, Koren S, Ghurye J, Rhie A, Phillippy AM, Rosen BD,
789 Bickhart DM, Smith TPL, et al. 2019a. New insights into mammalian sex chromosome
790 structure and evolution using high-quality sequences from bovine X and Y
791 chromosomes. *BMC Genomics* **20**: 1–11.
- 792 Liu Z, Li C, Xu J, Lan Y, Liu H, Li X, Maire P, Wang X, Jiang R. 2019b. Crucial and
793 Overlapping Roles of Six1 and Six2 in Craniofacial Development. *J Dent Res* **98**: 572–
794 579.
- 795 Mccullough DR, Jiang Z-G, Li C-W. 2009. Sika Deer in Mainland China.
- 796 Morin PA, Archer FI, Avila CD, Balacco JR, Bukhman Y V., Chow W, Fedrigo O,
797 Formenti G, Fronczek JA, Fungtammasan A, et al. 2021. Reference genome and
798 demographic history of the most endangered marine mammal, the vaquita. *Mol Ecol*
799 *Resour* **21**: 1008–1020.
- 800 Pan X, Cai Y, Li Z, Chen X, Heller R, Wang N, Wang Y, Zhao C, Wang Y, Xu H, et al.

801 2021. Modes of genetic adaptations underlying functional innovations in the rumen. *Sci*
802 *China Life Sci* **64**: 1–21.

803 Passemiers A, Moreau Y, Raimondi D. 2022. Fast and accurate inference of gene
804 regulatory networks through robust precision matrix estimation. *Bioinformatics* **38**:
805 2802.

806 Pemberton J, Johnston SE, Fletcher TJ. 2021. The genome sequence of the red deer,
807 *Cervus elaphus* Linnaeus 1758. *Wellcome Open Res* **6**: 336.

808 Price JS, Allen S, Faucheux C, Althnaian T, Mount JG. 2005. Deer antlers: a zoological
809 curiosity or the key to understanding organ regeneration in mammals? *J Anat* **207**: 603.

810 Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell
811 CM, Hart J, Landrum MJ, McGarvey KM, et al. 2014. RefSeq: an update on mammalian
812 reference sequences. *Nucleic Acids Res* **42**.

813 Qin T, Zhang G, Zheng Y, Li S, Yuan Y, Li Q, Hu M, Si H, Wei G, Gao X, et al. 2023. A
814 population of stem cells with strong regenerative potential discovered in deer antlers.
815 *Science* **379**: 840–847.

816 Radi SH, Vemuri K, Martinez-Lomeli J, Sladek FM. 2023. HNF4 α isoforms: the
817 fraternal twin master regulators of liver function. *Front Endocrinol (Lausanne)* **14**:
818 1226173.

819 Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: Reference-free quality,
820 completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**: 1–27.

821 Sangwung P, Zhou G, Nayak L, Chan ER, Kumar S, Kang DW, Zhang R, Liao X, Lu Y,

- 822 Sugi K, et al. 2017. KLF2 and KLF4 control endothelial identity and vascular integrity.
823 *JCI Insight* **2**.
- 824 Segre JA, Bauer C, Fuchs E. 1999. Klf4 is a transcription factor required for establishing
825 the barrier function of the skin. *Nat Genet* **22**: 356–360.
- 826 Seo MS, Park SB, Choi SW, Kim JJ, Kim HS, Kang KS. 2014. Isolation and
827 characterization of antler-derived multipotent stem cells. *Cell Transplant* **23**: 831–843.
- 828 Setty M, Kisieliovas V, Levine J, Gayoso A, Mazutis L, Pe’er D. 2019. Characterization
829 of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology* 2019
830 *37:4* **37**: 451–460.
- 831 Sorkin M, Huber AK, Hwang C, Carson WF, Menon R, Li J, Vasquez K, Pagani C, Patel
832 N, Li S, et al. 2020. Regulation of heterotopic ossification by monocytes in a mouse
833 model of aberrant wound healing. *Nat Commun* **11**.
834 <https://pubmed.ncbi.nlm.nih.gov/32024825/> (Accessed October 16, 2023).
- 835 Talenti A, Powell J, Hemmink JD, Cook EAJ, Wragg D, Jayaraman S, Paxton E,
836 Ezeasor C, Obishakin ET, Agusi ER, et al. 2022. A cattle graph genome incorporating
837 global breed diversity. *Nat Commun* **13**.
- 838 Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements
839 in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**.
- 840 Taylor M V., Hughes SM. 2017. Mef2 and the skeletal muscle differentiation program.
841 *Semin Cell Dev Biol* **72**: 33–44.
- 842 Torres-Estay V, Carreño D V., San Francisco IF, Sotomayor P, Godoy AS, Smith GJ.

- 843 2015. Androgen receptor in human endothelial cells. *J Endocrinol* **224**: R131.
- 844 Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter M, Mieszkowska N, Hall
845 N, Holland P, Durbin R, Richards T, Kersey P, et al. 2023. MitoHiFi: a python pipeline
846 for mitochondrial genome assembly from PacBio high fidelity reads. *BMC*
847 *Bioinformatics* **24**: 1–13.
- 848 Valkovic AL, Bathgate RA, Samuel CS, Kocan M. 2019. Understanding relaxin
849 signalling at the cellular level. *Mol Cell Endocrinol* **487**: 24–33.
- 850 Wang D, Berg D, Ba H, Sun H, Wang Z, Li C. 2019a. Deer antler stem cells are a novel
851 type of cells that sustain full regeneration of a mammalian organ—deer antler. *Cell*
852 *Death Dis* **10**.
- 853 Wang F, Zhang C, Kwagh J, Strassle B, Li J, Huang M, Song Y, Lehman B, Westhouse
854 R, Palanisamy K, et al. 2021. TGFβ2 and TGFβ3 mediate appropriate context-dependent
855 phenotype of rat valvular interstitial cells. *iScience* **24**.
- 856 Wang W, Hu CK, Zeng A, Alegre D, Hu D, Gotting K, Granillo AO, Wang Y, Robb S,
857 Schnittker R, et al. 2020. Changes in regeneration-responsive enhancers shape
858 regenerative capacities in vertebrates. *Science (1979)* **369**.
- 859 Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, Zhao Y, Han J, Pan X, Zheng Z, et al.
860 2019b. Genetic basis of ruminant headgear and rapid antler regeneration. *Science (1979)*
861 **364**.
- 862 Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson
863 HA, Flicek P, et al. 2020. An improved pig reference genome sequence to enable pig

864 genetics and genomics research. *Gigascience* **9**.

865 Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G,
866 Kriventseva E V., Zdobnov EM. 2018. BUSCO applications from quality assessments to
867 gene prediction and phylogenomics. *Mol Biol Evol* **35**: 543–548.

868 Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: Large-scale single-cell gene expression
869 data analysis. *Genome Biol* **19**: 1–5.

870 Wolock SL, Lopez R, Klein AM. 2019. Scrublet: Computational Identification of Cell
871 Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**: 281-291.e9.

872 Xie Z, Bailey A, Kuleshov M V., Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A,
873 Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, et al. 2021. Gene Set Knowledge
874 Discovery with Enrichr. *Curr Protoc* **1**.

875 Xing X, Ai C, Wang T, Li Y, Liu H, Hu P, Wang G, Liu H, Wang H, Zhang R, et al.
876 2023. The First High-quality Reference Genome of Sika Deer Provides Insights into
877 High-tannin Adaptation. *Genomics Proteomics Bioinformatics* **21**: 203–215.

878 Yu F, Li F, Yu P, Zhou B, Ye L. 2022. Identification and characterization of NFATc1+
879 skeletal stem cells in bone regeneration. *Cell Rep* **41**.

880 Zhang R, Dong Y, Xing X. 2022. Comprehensive transcriptome analysis of sika deer
881 antler using PacBio and Illumina sequencing. *Sci Rep* **12**: 16161.

882 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C,
883 Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS).
884 *Genome Biol* **9**: 1–9.

885

886

887 **Figures legends**

888 Figure 1 Assembly and annotation of the sika deer genome.

889 **A** Workflow of phased genome assembly of sika deer. **B** Hi-C interaction heatmap of
890 phased pseudochromosomes of sika deer genome. **C** Bar plot of BUSCO evaluation of
891 published genome annotations of sika deer and Cetartiodactyla species in RefSeq. **D** UMAP
892 of cell atlas of antler regeneration based on our genome assembly and annotation. **E** Dot plot
893 of cell marker expression profiles in each cell type. Dot size represent the proportion of cells
894 expressing gene in a cluster.

895

896 Figure 2 Chromatin accessibility and gene expression landscape for different organs of sika
897 deer.

898 **A** Schematic drawing of the study design covering primary organs and key stages of
899 antler regeneration and collecting paired expression and chromatin accessibility data. **B, C**
900 Hierarchical clustering heatmap of gene expression and chromatin accessibility of sika deer
901 organs. **D** Hierarchical clustering heatmap of TF-binding motif match profile showing similar
902 motif profile among self-renewable tissues.

903

904 Figure 3 Hub TFs in antler regeneration.

905 **A** Hierarchical cluster heatmap of the top 5 enriched motifs for each stage in antler
906 regeneration. **B** Function and time course schema of top 10 TFs for each stage, suggesting 4
907 core TFs and the dynamic from cell stemness and stimulus response to development. **C**
908 Functional enrichment of combinatory regulons of hub TFs in each antler regeneration stage.
909 **D** Hierarchical structure of 10 hub TFs subnetwork in antler regeneration at *dac5* with similar
910 pattern with time course schema. **E** UMAP projection of PMC cell lineage from antler and
911 mouse digit tip distinguishes PMC of antler and mouse. Dash circle highlighted the shared
912 activated stem cell in antler and regenerative digit tip. **F** Higher *KLF4* and *MYC* expression in
913 PMC of deer antler than mouse digit tips. **G** Genome track of antler specific element nearby
914 *MYC* (up), sequence alignment (lower left) and luciferase assay (lower right) suggest cervid-
915 specific insertion has significantly increased the expression of *MYC* in antler.

916

917 Figure 4 cTOP models cell programs in antler regeneration.

918 **A** cTOP couples cell expression modules and TF combinatory modules under the PECA
919 regulatory network to model cellular regulatory programs. **B, C** UMAP of cell types and
920 highest CPR in cell atlas at *dac5*. **D** Functional enrichment for TGs of each CPR. **E** UMAP
921 of pseudotime trajectory of stromal cells at *dac5* by Monocle3. **F** cTOP -relative TFs (red) and
922 genes (black) expression dynamics across the osteochondrogenesis trajectory. **G** Network of
923 CPR4 suggests *PRDMI*, *FOSL1* and *NFATC1* as the key factors for stem cell activation in
924 antler regeneration. **H** Expression dynamics of TFs in CPR4 during antler regeneration

925 suggest their early function in antler regeneration. **I** Expression dynamic across differentiation
926 of *SP7* and *SIX2* showing their program specific expression. **G** Schematic regulatory model of
927 cellular programs in antler stem cell differentiation. Black arrows represent differentiation,
928 and gray dashed arrows represent regulation.

929

930

